A Measure to compare Matchings in Marriage Markets

Florian M. Biermann

The International School of Economics at Tbilisi State University (ISET) is supported by BP, the Government of Georgia, the Norwegian Ministry of Foreign Affairs, Higher Education Support Program of the Open Society Institute, the Swedish International Development Agency and the World Bank.

International School of Economics at Tbilisi State University
16 Zandukeli Street, Tbilisi 0108, Georgia
www.iset.ge e-mail: publications@iset.ge

A Measure to compare Matchings in Marriage Markets☆

Florian M. Biermann

International School of Economics at Tbilisi State University, Georgia

Abstract

In matching markets the number of blocking pairs is often used as a criterion to compare matchings. We argue that this criterion is lacking an economic interpretation: In many circumstances it will neither reflect the expected extent of partner changes, nor will it capture the satisfaction of the players with the matching. As an alternative, we set up two principles which single out a particularly "disruptive" subcollection of blocking pairs. We propose to take the cardinality of that subset as a measure to compare matchings. This cardinality has an economic interpretation: The subset is a justified objection against the given matching according to a bargaining set characterization of the set of stable matchings. We prove multiple properties relevant for a workable measure of comparison.

Keywords: Stable marriage problem, Matching, Blocking pair, Instability, Matching comparison, Decentralized market, Bargaining set

1. Introduction

While almost all matching theory only distinguishes between stable and unstable matchings, sometimes it is necessary to compare multiple unstable matchings.

For example, an experimental economist who simulates a matching market

☆The author is indebted to Naomi Hass, Todd Kaplan, Flip Klijn, Bezalel Peleg, Andrés Perea, Amnon Schreiber, and Eyal Winter for numerous useful comments and pieces of advice.

Email address: `f.biermann@iset.ge` (Florian M. Biermann)

URL: `http://florianbiermann.net` (Florian M. Biermann)

in the classroom or in a computer laboratory may find that some of the experiments conducted did not result in stable matchings. In order to interpret the results, a comparison could be made regarding "how close" they are to stability. This problem was encountered in Niederle and Roth (2007), and the authors solved it by taking the number of blocking pairs as a criterion to compare different matchings.

A related problem occured in Roth and Xing (1997), where the matchings resulting from a simulation were not generally stable, yet had to be compared with each other. In a similar manner, the number of blocking agents, i.e. those players who are part of at least one blocking pair, was taken as a criterion of comparison.[1,2]

Another necessity for the comparison of unstable matchings emerges when the matchmaker is not trying to achieve stability, but instead pursues a different goal. If the objective does not single out a unique matching, then one must be selected from those which fulfill the primary requirement. Typically, such a situation occurs when the matchmaker primarily wants to maximize the number of matched pairs, i.e. wants to find a *maximal* matching. This is a reasonable objective for numerous markets in which the social benefit or the matchmaker's profit hinges on the number of matched players, while stability is deemed not so important. Biró et al. (2010) describe many such situations and mention the related literature. One of their examples is an organ exchange market, where the maximality of the matching is the primary goal: The size of the matching determines the number of transplantations, which have the potential to be life saving. Moreover, blocking pairs existing in the final matching will not cause further reshuffling as the agents will usually not undergo additional operations just to resolve blocking pairs. Yet the *satisfaction* the players obtain from the final matching is of utmost importance:

[1] See in particular pp. 318-320 in their article.

[2] Roth and Xing (1997) analyze the entry-level labor market for clinical psychologists, which they do not model as a marriage market, but as a many-to-one matching market. It is well known that many results derived for marriage markets carry over to many-to-one matching problems (cf. Roth and Sotomayor (1990), chapter 5). Notably, the problems associated with counting blocking pairs, which motivate this paper, exist in the same way in many-to-one models. Therefore the reasoning presented here for marriage markets applies in the same way to many-to-one models, and the measure of instability advocated here can be naturally adapted to a many-to-one framework. For the conceptual purposes of this paper it is unnecessary to cope with the considerably higher complexity of many-to-one models.

Centralized matching regimes which are not accepted by the agents usually get undermined.[3] The more agents are not satisfied with the matchmaker's performance, the higher is the risk that the reputation of the centralized system will deteriorate, causing players to search for partners in decentralized ways.
Other applications mentioned in said article are school placement, assigning students to university projects, a certain bipartite matching problem of the US Navy, and even the optimal pairing of players in chess tournaments. For these situations, in which stability is not the most important concern, Hamada et al. (2009) and Biró et al. (2010) develop algorithms which create maximal matchings with the least number of blocking pairs.[4]
Finally, there are situations in which stability cannot be achieved for exogenous reasons, and in the absence of a stable matching the matchmaker has to rank different unstable outcomes. For example, Khuller et al. (1994) develop an "online" matching algorithm for a situation in which all women are in the market from the start, while the men enter sequentially. As soon as a man has entered the market, the algorithm must match him to a woman immediately. No assignments once made can be undone on later stages. Obviously, no algorithm can guarantee that the outcome is stable. Thus Khuller et al. (1994) design their algorithm so as to minimize the expected number of blocking pairs for the resulting matching.[5]

Indeed, counting blocking pairs and taking their number as a criterion to compare matchings makes a lot of sense at first sight.[6] A blocking pair is

[3]Undermining occurs for example through the phenomenon called "unraveling" (cf. Niederle and Roth (2003)). Also, the history of the NIMP algorithm provides evidence that acceptance by the participants is essential for the survival of a matching regime (cf. Roth (1984)). Similar observations were made in entry-level labor markets for physicians in the U.K. (cf. Roth (1991)).

[4]The paper by Hamada et al. (2009) builds on Biró et al. (2010). The latter was already published preliminarily as a conference proceeding and as a working paper in 2008.

[5]Real-world matching markets in which stability is *practically* unattainable are prevalent. Uncontrolled influx and outflux of market participants, as modeled by Khuller et al. (1994), can frequently be observed. In other (decentralized) markets, information deficits may be the main factor for the absence of stable matchings (cf. Eriksson and Häggström (2008)).

[6]Without changing the concept qualitatively, one may also divide the number of blocking pairs which exist for a matching by the number of all possible pairs, as advocated by Eriksson and Häggström (2008). This procedure is only necessary, however, if one wants

a source of dissatisfaction on the part of its members. Therefore a matchmaker might lose customers and profits if too many players eventually find themselves in blocking pairs. Moreover, one might consider the number of blocking pairs as a proxy for the amount of partner changes imminent at a given state of the market. This interpretation also corresponds in a direct way with the term “measure of instability”, which is used occasionally in the literature.

Despite the superficial reasonability of counting blocking pairs, we think that the concept is problematic. Often there is no connection between the satisfaction of the players with a certain matching and the number of blocking pairs. Consider the following example: let μ be a matching and let $\mathscr{B}(\mu)$ be the collection of blocking pairs for μ. If $\mathscr{B}(\mu)$ is no matching, these blocking pairs cannot be satisfied simultaneously.[7] In the most extreme case, a given player $m \in M$ (or $w \in W$) is member of *all* pairs in $\mathscr{B}(\mu)$. Of these blocking pairs, only one can be resolved. It seems questionable whether such blocking pairs would generate the same level of dissatisfaction as an equal number of blocking pairs which actually could be satisfied simultaneously. In the outlined situation, there may be n different women $\{w_1, \ldots, w_n\}$ who form blocking pairs with m, but each of these women knows that m, if he could decide which of the blocking pairs was to be satisfied, would marry $w := \max_{\succ_m}\{w_1, \ldots, w_n\}$. So if the dissatisfaction with the matching is based on rational considerations, essentially only one woman and one man would be discontent with the matching μ, namely w and m.[8]

If the set of blocking pairs whose cardinality is counted was a matching, a feature of the concept we propose, this problem would entirely disappear.

to compare matchings from different markets. If the matchings to be compared are within the same market, as will be assumed throughout this paper, one does not lose anything by taking the absolute number of blocking pairs.

[7]In this paper a matching is defined as a collection of pairs such that none of them share a player, see definition 2 on page 7. This definition was used before (for example in Blum and Rothblum (2002)) and it is equivalent with the standard function definition in Roth and Sotomayor (1990).

[8]Our point may be further illustrated by the following example: A single super model forms blocking pairs with thousands of men in a population. But do these blocking pairs cause discomfort among rational men? We do not believe so. A rational man knows that the supermodel will by all likelihood not consider him to be the most attractive partner among those with whom she forms blocking pairs. The blocking pair in which he participates is rather fantasy than a real option.

Then the pairs contained in that set could be satisfied in parallel, and all of these blocking pairs would be forgone opportunities to improve the outcomes of the participating players. These improvements could have materialized by these players if they would not have participated in the centralized matching mechanism, causing justified dissatisfaction with the matchmaker.
Yet it is equally problematic to take the number of blocking pairs as a measure of the "degree of instability" of a matching. Again, n blocking pairs which have a player in common cannot be satisfied simultaneously, hence only one of the n pairs could actually trigger a partner change. For this reason, the instability of matchings which have a high number of "hypothetical" blocking pairs (blocking pairs which share players with other blocking pairs) could be overstated by this measure. At the same time, one can construct examples in which even one single blocking pair may trigger off a vacancy chain changing the assignments of *all* players in the market. The instability of a matching with few but highly disruptive blocking pairs can be drastically understated by the traditional measure. The connection between the "degree of instability" of a matching and its number of blocking pairs is very lose at best.[9]

This paper offers an alternative measure for comparing matchings. We believe that certain subsets of the blocking pairs have a particular significance, both regarding their "disruptive potential" as well as in terms of player satisfaction. We call them *permissible sets* of blocking pairs for a matching. It will be shown in proposition 2 on page 12 that all permissible sets of blocking pairs for a matching have the same cardinality. So their cardinality can be used as a measure to compare matchings. A matching which has a higher value according to this measure is expected to show more reshuffling as well as higher dissatisfaction among the players.
Moreover, we claim that a permissible set of blocking pairs constitutes not only a *possible* transformation of the market, but also a *likely* transformation among shortsighted players. If $D(\mu)$ is a permissible set of blocking pairs for

[9] In a model with undisclosed preferences which get disclosed upon random encounters of the players, the total amount of blocking pairs would be an indicator for the expected readjustments within a certain time span or in a given amount of stages. This would be an interesting model, but it would go beyond the setting discussed here. Here we keep to the assumption that the existence of a blocking pair is known to the players who form it, as it is standard in most of matching theory.

a matching μ, then our claim is based on the fact that $D(\mu)$ can be interpreted as a *justified objection* against μ according to a bargaining set which has the appealing property that it coincides with the set of stable matchings. So if one believes in the empirical relevance of bargaining set concepts and Gale-Shapley stability, one can conjecture that the blocking pairs which get satisfied at an unstable matching μ comprise a permissible set.[10]
Unfortunately, even if the set $D(\mu)$ of blocking pairs to be counted comprises a justified objection against μ, its size is just an indicator for the first-order dynamics emerging from the given matching. The concept proposed here says nothing about further market transformations which could take place after the counted blocking pairs were resolved.[11] Therefore the use of the concept presented here as a "measure of instability" is limited. Yet it improves on the simple counting of blocking pairs, which cannot even predict the first transformation of the market and does not allow for inference on the n^{th}-order dynamics either. Furthermore, the limitation may not be so severe if divorces are costly, as it is the case in many practical applications. The more unattractive partner changes become, the more important becomes the first transformation of the market, while multistage dynamics will be shorter and less likely to occur. The example of an organ exchange market, in which people refrain from undergoing further surgeries in order to resolve blocking pairs, was already mentioned.[12]

At the core of the approach introduced here is the selection of a matching to be formed from the set $\mathscr{B}(\mu)$ of blocking pairs for a matching μ. This is not trivial, as usually many matchings can be formed from the elements in $\mathscr{B}(\mu)$. We offer a rule for this selection. It will be stated and formalized in Section 3, and the resulting *permissible sets* of blocking pairs will be interpreted economically in Section 4. Furthermore, these sets will be shown to have interesting and useful features in Sections 5, 7, and 8. The most important of these features is the fact that all permissible sets of blocking

[10]More on empirical support for bargaining sets can be found in footnote 19 on page 11.

[11]The issue of n^{th}-order dynamics emerging from a matching is briefly discussed in Appendix B on page 26.

[12]Another example provides Roth (1991): Many entry-level markets for physicians in the U.K. were so small that most of the players in the market knew each other personally. As players were expected to accept the matching determined by the centralized matching algorithm, it could be costly in terms of reputation to resolve blocking pairs.

pairs have the same cardinality for a given matching, so that this cardinality constitutes a well-defined measure of comparison. Section 6 shows *how* to find the permissible sets of blocking pairs for a given matching. Section 9 illuminates a connection between the measure put forward here and the total number of blocking pairs used previously. The paper is concluded with Section 10. In the Appendix B we look at n^{th}-order dynamics which can be derived from the concept of permissible sets.

2. Preliminaries

Definition 1 (Marriage Market). *A marriage market is a triplet $(M, W, \boldsymbol{\succeq})$, where M and W are disjoint finite sets and $\boldsymbol{\succeq}$ is a set which contains for each $m \in M$ a linear order $\succeq_m$ defined over the set $\{m\} \cup W$.*[13] *In the same way, $\boldsymbol{\succeq}$ contains for each $w \in W$ a linear order $\succeq_w$ defined over the set $\{w\} \cup M$.*

We refer to $\boldsymbol{\succeq}$ as a *preference profile.* The item m, over which the preference order $\succeq_m$ is defined, stands for m's option of being single. Likewise the item w, over which the preference order $\succeq_w$ is defined, stands for w's option of being single. For $x \in M \cup W$, the strict[14] order $\succ_x$ is derived from $\succeq_x$ by the rule $a \succ_x b \Leftrightarrow a \succeq_x b \ \wedge \ a \neq b$.

As in Blum and Rothblum (2002), we define a matching to be a set of pairs, which is equivalent to the usual function definition of matchings:

Definition 2 (Matching). *A matching in the marriage market $(M, W, \boldsymbol{\succeq})$ is a set $\mu \subseteq M \times W$ such that if $(\hat{m}, \hat{w}), (\bar{m}, \bar{w}) \in \mu$, then $\hat{w} = \bar{w}$ if and only if $\hat{m} = \bar{m}$.*

If for $m \in M$ there exists no $w \in W$ with $(m, w) \in \mu$, then we say that m is *single* under matching μ. Correspondingly, if for $w \in W$ there exists no $m \in M$ with $(m, w) \in \mu$, then we say that w is *single* under matching μ.
To ease notation, for a pair $(m, w) \in \mu$ we will write $\mu(m)$ to denote m's partner under μ, i.e. $\mu(m) := w$. In this case, we will also write $\mu(w)$ to

[13]Linearity of an order means that it fulfills antisymmetry, transitivity, and totality. Due to antisymmetry, a linear order does not allow for ties between unequal elements: For $z, y \in \{m\} \cup W$ with $z \neq y$ either holds $z \succeq_m y$ or $y \succeq_m z$, but not both.

[14]An order is strict if it fulfills irreflexivity, asymmetry, and transitivity.

denote w's partner under μ, i.e. $\mu(w) := m$. If there is *no* pair in μ of which a player $x \in M \cup W$ is a member, then we denote by $\mu(x)$ the player x himself, i.e. $\mu(x) := x$.

Definition 3 (Blocking pair). *Let μ be a matching. A pair $(m, w) \in M \times W$ with $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$ is a blocking pair for the matching μ.*

The set of blocking pairs which exist for a matching μ is denoted by $\mathscr{B}(\mu)$.

Definition 4 (Individual Rationality). *A matching μ is individually rational if for every player $x \in M \cup W$ holds $\mu(x) \succeq_x x$.*

Definition 5 (Stability, Gale and Shapley (1962)). *A matching μ is stable if it is individually rational and no blocking pairs exist for μ.*

The following notation will be used throughout the paper: For $U \subseteq M \times W$ we denote by $p(U) \subseteq M \cup W$ the set of those players who are member of a pair in U, formally

$$p(U) = \{x \in M \cup W \mid \exists y \in M \cup W : \ (x, y) \in U \vee (y, x) \in U\}.$$

3. Permissible collections of blocking pairs

We denote by $D(\mu) \subseteq \mathscr{B}(\mu)$ that set of blocking pairs whose cardinality we propose to take as a measure of comparison. As mentioned in the introduction, other authors usually set $D(\mu) := \mathscr{B}(\mu)$. $D(\mu)$ can only be a possible deviation from the matching μ if $D(\mu)$ is a matching, i.e. $D(\mu)$ is a collection of such blocking pairs which can get satisfied simultaneously. This requirement is stated as:

Principle 1. *The set $D(\mu)$ is a matching.*

Of course there are many subsets of $\mathscr{B}(\mu)$ which are matchings, so Principle 1 alone does not give an answer which blocking pairs should be counted. Our proposal for solving this problem makes use of the concept of domination, brought forward in Klijn and Massó (2003):

Definition 6. *Let $(\hat{m}, \hat{w}), (m, w) \in M \times W$. Then $(\hat{m}, \hat{w})$* dominates *(m, w) if one of the following two conditions is fulfilled: 1) $\hat{m} = m$ and $\hat{w} \succ_m w$ or 2) $\hat{w} = w$ and $\hat{m} \succ_w m$.*

For $(m,w) \in M \times W$, denote by $dom((m,w)) \subseteq M \times W$ the set of pairs which are dominated by (m,w). Note that the irreflexivity of the relations $\succ_m$ and $\succ_w$ ensures $(m,w) \notin dom((m,w))$, $\forall (m,w) \in M \times W$. The following simple lemma on domination is used later:

Lemma 1. *Let (m,w) and $(\hat{m},\hat{w})$ be two different blocking pairs in $M \times W$. If $m = \hat{m}$ or $w = \hat{w}$, then exactly one of the following statements is true:*

1. $(m,w) \in dom((\hat{m},\hat{w}))$
2. $(\hat{m},\hat{w}) \in dom((m,w))$.

The proof of this lemma is provided in the appendix (page 24). Using domination, we state

Principle 2. *If $(m,w) \in \mathscr{B}(\mu) \backslash D(\mu)$, then $(m,w) \in dom((\hat{m},\hat{w}))$ for some $(\hat{m},\hat{w})$ in $D(\mu)$.*

Verbally, a blocking pair (m,w) will not be counted only if another blocking pair $(\hat{m},\hat{w})$ which dominates (m,w) *will* be counted. From an economic viewpoint, the formation of $(\hat{m},\hat{w})$ prevails over the formation of (m,w), because $(\hat{m},\hat{w})$ and (m,w) share a member who prefers his or her partner in $(\hat{m},\hat{w})$ and thus would refrain from entering (m,w).[15]
A set of blocking pairs which fulfills the Principles 1 and 2 is referred to as *permissible*. In Section 7 will be shown that a permissible set of blocking pairs for a matching μ is empty if and only if μ is a stable matching.

4. The economic basis of permissible sets of blocking pairs

In this section, we provide an economic motivation for the permissible sets of blocking pairs defined above. Klijn and Massó (2003) adapt the bargaining set of Zhou (1994) to marriage markets and prove that it coincides with the set of *weakly stable*[16] matchings.[17] Here we will take a different direction:

[15]For another economic interpretation of *domination*, drawing on the farsightedness of the players, see Klijn and Massó (2003) p. 94.

[16] If all blocking pairs for a matching are dominated by other blocking pairs, then a matching is called weakly stable. Klijn and Massó (2003) show by example that the set of weakly stable matchings may be a superset of the set of stable matchings, i.e. there are marriage markets in which a matching μ is not stable, but for any $(m,w) \in \mathscr{B}(\mu)$ exists a pair $(\bar{m},\bar{w}) \in \mathscr{B}(\mu)$ with $(m,w) \in dom((\bar{m},\bar{w}))$.

[17]The Zhou bargaining set is obtained from the bargaining set of Mas-Colell (1989) by replacing a weak inequality by a strict inequality in the definition of the objection

We will try to find an economically reasonable definition of a bargaining set for marriage markets which coincides with the set of *stable* matchings. Why do we need such a bargaining set?
Our goal is to interpret permissible sets as justified objections of the participating players against the given matching. Because the cardinality of the permissible sets is our measure of matching comparison, it should be 0 if a matching is stable. This is an indispensable condition for the claim that our measure indicates the "degree of instability" of a matching in some sense. On the other hand, if there *are* blocking pairs for a matching, it is unstable, and thus our measure should *not* assume the value 0. In this case, there should be a nonempty permissible set and thus a nonempty justified objection against that matching. For these reasons we are looking for a bargaining set with two properties:

1. There should be no justified objection against a matching if and only if it is stable.
2. The justified objections against a matching should be the permissible sets.

Obviously, the bargaining set of Zhou (1994) and its adaptation by Klijn and Massó (2003) does not fulfill our demands: There can be unstable matchings which are weakly stable, and which are therefore in the bargaining set of Klijn and Massó (2003). So according to their concept, there exist no justified objections against these unstable matchings. In contrast, the following definition of a bargaining set has the desired features:

Definition 7 (Objection). *An* objection *against a matching* μ *is a matching* $S \neq \emptyset, S \subseteq \mathscr{B}(\mu)$.

Definition 8 (Counterobjection). *A* counterobjection *against an objection* S *is a matching* $T \neq \emptyset, T \subseteq \mathscr{B}(\mu)$, $T \nsubseteq S$, *such that for any pair* $(m, w) \in T$ *and any pair* $(\hat{m}, \hat{w}) \in S$ *holds* $(m, w) \notin dom((\hat{m}, \hat{w}))$.

As it is known from other bargaining set concepts, an objection for which no counterobjection exists is called *justified*.

and imposing further restrictions on the counterobjection. Both bargaining sets have interesting mathematical properties, a discussion of whom can be found in Peleg and Sudhölter (2007).

In short, an objection $S \subseteq \mathscr{B}(\mu)$ against a matching μ is a matching formed from blocking pairs for μ. A counterobjection $T \subseteq \mathscr{B}(\mu)$ is a matching which is also formed from blocking pairs for μ and it consists only of pairs which are not dominated by pairs of the objection. This is economically reasonable: If a pair $(m, w) \in T$ *was* dominated by a pair in S, then either m or w would strictly prefer to keep to the objection and the counterobjection could not form. If, on the other hand, all pairs in T were *not* dominated by pairs in S and hence T was a valid counterobjection, then the supporters of S would have no arguments against the formation of T. They would not find a player participating in T whom they could convince to stay in S – all players participating in T would weakly prefer T over S.
In accordance with economic intuition, the condition $T \nsubseteq S$ rules out the possibility that an objection can be countered by itself or by a subset of itself, because in such a situation the counterobjection would yield exactly the same payoff as the objection to all of its participants.[18]

The proof that there exists no justified objection against a matching μ if and only if μ is stable is provided in Section 7. The following result validates the interpretation of a permissible set $D(\mu)$ as a justified objection of a group of players against the matching μ.

Proposition 1. *$D(\mu)$ is a permissible set of blocking pairs if and only if it is a justified objection against μ.*

The proof, which has no aesthetic value, is given in the appendix on page 25. In view of the preceding proposition, it becomes clear that our measure is the cardinality of groups of players who can come together to improve their outcome independently of the other players. In this way, they form an objection against a matching. But among such coalitions, only those are considered which cannot be blocked by counterobjections. If one accepts that in general bargaining concepts have real world significance, it makes sense to attribute a strong potential to reshuffle a matching market to those coalitions which are justified objections.[19] Moreover, as this bargaining set coincides with the

[18]Remind that S and T are collections of pairs, not players. As the utility of a player is solely determined by the pair of which he or she is a member, no player in T would have a gain from forming a counterobjection T against S if $T \subseteq S$.

[19]Section 11 of Maschler (1992) reviews empirical evidence for and against bargaining set concepts and discusses its validity. The data on which Maschler bases his analysis was both

set of stable matchings, the conjecture that its justified objections play a significant role in real world marriage markets is indirectly supported by the undoubted empirical relevance of Gale-Shapley stability.
Furthermore, dissatisfaction will particularly prevail among players who find themselves in a justified objection (and not just in a blocking pair). If a player realizes that he or she is member of a justified objection against a matching, it does not only mean that an improvement for himself or herself was left out by the matchmaker. Aggravating would be the fact that without counterobjections, the improvement would be practically attainable through decentralized negotiations between the players.
Finally, Ehlers (2007) adapted Von Neumann-Morgenstern stable sets to marriage markets. A set of matchings is a von Neumann-Morgenstern stable set if it is *internally stable* and *externally stable* (for details see Ehlers (2007)). Without proof, we note that the permissible sets of blocking pairs for a matching μ comprise an internally stable set of matchings.

5. A workable measure to compare matchings

To evaluate the next result correctly, it is important to remind that by definition, *any* matching which is a subset of $\mathscr{B}(\mu)$ is an objection. Moreover, even if for some reason all objections were formed by the same set of players, from the above definition does not follow that the counterobjections are formed by the same set of players: If a counterobjection T against an objection S contains a pair which is an element of S, it can be included in T or left out without changing the fact that T is a counterobjection against S. Considering these facts, it is somewhat surprising that every *justified objection* against a given matching μ is formed by the same set of players, as will be shown next.

Proposition 2. *Let S and T be justified objections against an individually rational matching μ. Then $p(S) = p(T)$.*[20]

generated in laboratory experiments (pp. 638-641) and obtained from real world situations (pp. 641-642). In 1992 the empirical foundation of bargaining sets was controversial but encouraging. Maschler himself assumes a rather critical standpoint, while authors he cites, notably Kahan and Rapoport (1984), are very positive about the empirical importance of bargaining sets.

[20] For the notation $p(U)$, see page 8.

Proof. S and T are matchings, so for notational consistency we set $\mu_S := S$ and $\mu_T := T$. Without loss of generality, assume there is a player $\bar{m} \in M$ with $\bar{m} \in p(S) \backslash p(T)$. We will now derive a contradiction from this assumption. First of all, if $\mu_S(\bar{m}) \in p(S) \backslash p(T)$, then there is no pair in T which shares a member with $(\bar{m}, \mu_S(\bar{m}))$. Consequently, there is no pair $(\hat{m}, \hat{w}) \in T$ with $(\bar{m}, \mu_S(\bar{m})) \in dom((\hat{m}, \hat{w}))$ and so $(\bar{m}, \mu_S(\bar{m}))$ is a counterobjection against T, contradicting our assumption that T is a justified objection. So it must hold $\mu_S(\bar{m}) \in p(T)$.

Set $K := \{ \; x \in M \cup W \mid \mu_S(x) \neq \mu_T(x) \; \}$, i.e. K is the set of players who do not have the same partners under μ_S and μ_T. Note that $\bar{m} \in K$. Now we employ a graph-theoretic argument. Construct a graph $\mathscr{G} = (K, E)$ whose vertices are the elements in the set K. Let there be an edge in E between $m \in K$ and $w \in K$ if $(m, w) \in \mu_S \cup \mu_T$.[21] In this graph, players in K who are single neither under matching μ_S nor under matching μ_T are members of two different pairs in $\mu_S \cup \mu_T$, hence they have a degree of 2. Players in K who are single under exactly one of both matchings are member of exactly one pair in $\mu_S \cup \mu_T$, hence they have a degree of 1. (Remind that players who are singles under both μ_S and μ_T and players who have the same partners under both matchings are not members of K, hence they are not vertices of the graph $\mathscr{G}$.) Because all vertices in $\mathscr{G}$ have degree of either 1 or 2, all connected components of the graph $\mathscr{G}$ must be circuits or simple chains.[22]

If a component of $\mathscr{G}$ is a circuit, all vertices in that component have degree 2. From $\bar{m} \notin p(T)$ follows that $\bar{m}$ has a degree of 1, so the connected component of which $\bar{m}$ is a node must be a simple chain. The proof will be completed by showing that this simple chain cannot be a component of the graph $\mathscr{G}$.[23]

[21] This kind of graph is popular in matching theoretic proofs: It is similar to the so called bi-choice graph introduced by Klaus and Klijn (2010). Like the bi-choice (di)graph, which helps its inventors to prove and reprove a couple of results on roommate problems in Klaus and Klijn (2010) and Klaus et al. (2009), the graph defined here connects a player $x \in K$ with the partners x has under two different matchings (in Klaus et al. (forthcoming) the bi-choice graph is not used anymore). This feature is also shared by the graph constructed in "Case 1" of the proof of Theorem 1 in Cechlárová (2002).

[22] A simple chain in a graph is a sequence of *distinct* vertices $v_1, \ldots, v_n$ with an edge between each v_i and v_{i+1} for $1 \leq i \leq n-1$. A circuit is a simple chain $v_1, \ldots, v_n$ with an additional edge between v_1 and v_n, and all edges which are part of the circuit are *distinct* (cf. Roberts and Tesman (2009), p. 135).

[23] For the proof, one could also use the bi-choice (di-)graph of Klaus and Klijn (2010). Therefore one would first have to show that in the bi-choice digraph of the matchings μ_S

By contradiction, assume $v_1, \ldots, v_n$ to be the simple chain of of players starting with $\bar{m}$ (i.e. $v_1 = \bar{m}$). From $(v_1, v_2) \in \mu_S$ follows inductively:

$$\begin{aligned} \forall \;\; 1 \leq j \leq n-1: & \\ & j \text{ is odd } \Rightarrow (v_j, v_{j+1}) \in \mu_S, \\ & j \text{ is even } \Rightarrow (v_{j+1}, v_j) \in \mu_T. \end{aligned} \tag{1}$$

If we walk along the simple chain $v_1, \ldots, v_n$, we recognize another rule:[24]

$$\begin{aligned} \forall \;\; 1 \leq j \leq n: & \\ & j \text{ is odd } \Rightarrow \mu_S(v_j) \succ_{v_j} \mu_T(v_j), \\ & j \text{ is even } \Rightarrow \mu_T(v_j) \succ_{v_j} \mu_S(v_j). \end{aligned} \tag{2}$$

This is shown by induction: Obviously, (2) holds for v_1, because m is single under μ_T but not under μ_S. As μ_S is an individually rational matching, m prefers his partner under μ_S over being single under matching μ_T. By contradiction, assume (2) would not be true for all j, and let $1 \leq j \leq n$ be the lowest integer such that (2) is *not* fulfilled. Furthermore, assume j to be odd (for an even j, the argument is symmetrical). Then for v_j must hold $\mu_S(v_j) \prec_{v_j} \mu_T(v_j)$, while according to (2) for v_{j-1} holds $\mu_T(v_{j-1}) \succ_{v_{j-1}} \mu_S(v_{j-1})$. From (1) follows $(v_j, v_{j-1}) \in \mu_T \subseteq \mathscr{B}(\mu)$, and both v_j and v_{j-1} prefer each other over their partners under μ_S. Because in K are only players who have different partners under μ_S and μ_T, it is ensured that $(v_j, v_{j-1}) \notin \mu_S$. Although $(v_j, v_{j-1}) \in \mathscr{B}(\mu) \backslash \mu_S$, there is no pair in μ_S which dominates (v_j, v_{j-1}), contradicting our assumption that μ_S was a justified objection against μ. This proves the correctness of (2).
Now consider player v_n: If v_n is odd, then by (1) $(v_n, v_{n-1}) \in \mu_T$, while he is single under μ_S (otherwise he would not be the last element of the simple chain). So individual rationality of μ_T implies $\mu_T(v_n) \succ_{v_n} \mu_S(v_n)$, contradicting (2). In the same way, if v_n is even, then by (1) $(v_{n-1}, v_n) \in \mu_S$, while she is single under μ_T (otherwise she would not be the last element of the simple chain). So individual rationality of μ_S implies $\mu_S(v_n) \succ_{v_n} \mu_T(v_n)$,

---

and μ_T, no two players point at each other. This is a requirement needed for Lemma 1 of Klaus et al. (2009) to hold. Afterwards, from their Lemma 1 follows that there are only cycles and loops in the bi-choice graph, implying that there is no simple chain.

[24] In the context of roommate problems, rules (1) and (2) were also derived in Step 9 of the proof of Theorem 1 in Diamantoudi et al. (2004) for a similar graph.

again contradicting (2). If follows that there can be no component of the graph $\mathscr{G}$ which is a simple chain starting with $\bar{m} \in p(S) \backslash p(T)$, implying that there exists no player $\bar{m} \in p(S) \backslash p(T)$. $\square$

Herewith it is proved that for a matching μ, all sets $D(\mu)$ have the same cardinality. So this number is well-defined and can be used as a measure of comparison.

6. Finding all permissible sets of blocking pairs for a given matching $\boldsymbol{\mu}$

Only if at least one permissible set of blocking pairs for a matching μ is identified, its elements can be counted. For a concept with practical aspirations, it is therefore essential to show how to *find* a permissible set of blocking pairs for an arbitrary unstable matching μ. Beyond that, further analysis of the concept may make it necessary not only to identify one permissible set, but to find all of them. Fortunately, there is a simple way to achieve this goal.
Let μ be an individually rational matching in the market $(M, W, \succeq)$. For $x \in M \cup W$ we denote by $B_\mu(x) \subseteq M \cup W$ the set

$$B_\mu(x) := \{y \in M \cup W \mid (x, y) \vee (y, x) \in \mathscr{B}(\mu)\},$$

i.e. the set of all players with whom x forms a blocking pair for μ. Furthermore, for any $x \in M \cup W$ we consider a preference order $\tilde{\succeq}_x$ which is defined on the same domain as $\succeq_x$. $\tilde{\succeq}_x$ has the following properties:[25]

$$z, y \in B_\mu(x) : z \tilde{\succ}_x y \Leftrightarrow z \succ_x y \tag{3}$$

and

$$z \notin B_\mu(x) \Leftrightarrow x \tilde{\succ}_x z. \tag{4}$$

Those comparisons not determined by the above rules must be chosen arbitrarily subject to transitivity and antisymmetry of the resulting order. Given a market $(M, W, \succeq)$ and an individually rational matching μ in this market, a preference order $\tilde{\succeq}_x$ with the above two properties exists for any $x \in M \cup W$. This can be seen as follows: Clearly, the preferences $\succeq_x$ for $x \in M \cup W$

[25] As before, the relation $\tilde{\succ}_x$ is derived from $\tilde{\succeq}_x$ by the rule $a \tilde{\succ}_x b \Leftrightarrow a \tilde{\succeq}_x b \ \wedge \ a \neq b$.

already fulfill (3). In two steps we can manipulate $\succeq_x$ to make it compatible with (4). At first, we rank the single option x directly below the element $\min_{\succeq_x} B_\mu(x)$ and denote the resulting preference order by $\succeq'_x$. Afterwards, we assign to any element $y \notin B_\mu(x)$ with $y \succ'_x x$ an arbitrary rank below x (avoiding ties). The resulting preference order fulfills (3) and (4). Furthermore, it is transitive, total, and antisymmetric, because in our manipulation we did not introduce ties into $\tilde{\succeq}_x$ and so the linearity of $\succeq_x$ carries over to $\tilde{\succeq}_x$. We need the following lemma:

Lemma 2. *Let μ be a matching in a market $(M, W, \succeq)$ and let $(m, w) \in M \times W$ be an individually rational pair*[26] *with $(m, w) \notin \mu$. If there is no $(\hat{m}, \hat{w}) \in \mu$ with $(m, w) \in dom(\hat{m}, \hat{w})$, then (m, w) is a blocking pair for μ.*

The proof is stated in the appendix on page 25.

Now let μ be an unstable but individually rational matching in the market $(M, W, \succeq)$, and let $\tilde{\succeq}$ be a preference profile which is constructed according to the conditions (3) and (4) based on the sets $B_\mu(x)$ for all $x \in M \cup W$. With these definitions, we can state:

Proposition 3. *A set $D \subseteq \mathscr{B}(\mu)$ is a permissible set of blocking pairs for the unstable but individually rational matching μ in the market $(M, W, \succeq)$ if and only if D is a stable matching in the market $(M, W, \tilde{\succeq})$.*

Proof. Form a stable matching $\tilde{\mu}$ in the market $(M, W, \tilde{\succeq})$. For $x \in M \cup W$, the players preferred over the single option according to the preferences $\tilde{\succ}_x$ are those with whom x forms a blocking pair for μ. So from the fact that $\tilde{\mu}$ is individually rational under preferences $\tilde{\succeq}$ follows $\tilde{\mu} \subseteq \mathscr{B}(\mu)$.[27] Moreover, $\tilde{\mu} \neq \emptyset$, because $\mathscr{B}(\mu) \neq \emptyset$ and for any pair $(m, w) \in \mathscr{B}(\mu)$ holds $m \tilde{\succ}_w w$ and $w \tilde{\succ}_m m$ by condition (4) above. So if $\tilde{\mu}$ was the empty matching, then a pair $(m, w) \in \mathscr{B}(\mu)$ would be a blocking pair for $\tilde{\mu}$ under preferences $\tilde{\succeq}$, conflicting with $\tilde{\mu}$'s stability in the market $(M, W, \tilde{\succeq})$. With $\tilde{\mu} \subseteq \mathscr{B}(\mu)$ and $\tilde{\mu} \neq \emptyset$ we have established that the matching $\tilde{\mu}$ is an objection against μ in the market $(M, W, \succeq)$. Next will be shown that $\tilde{\mu}$ is justified. Assume by

[26] This means that both $w \succ_m m$ and $m \succ_w w$ are fulfilled. Klaus et al. (2009) would say that m and w are *not matched in an individually irrational way* (see the proof of their Theorem 1).

[27] As before, the set $\mathscr{B}(\mu)$ is the set of blocking pairs for μ in the market $(M, W, \succeq)$.

contradiction there would be a counterobjection T against $\tilde{\mu}$. Then because $T \nsubseteq \tilde{\mu}$, there exists a pair $(m,w) \in T, (m,w) \notin \tilde{\mu}$, such that for no pair $(\bar{m},\bar{w}) \in \tilde{\mu}$ holds $(m,w) \in dom((\bar{m},\bar{w}))$. From $(m,w) \in T \subseteq \mathscr{B}(\mu)$ follows that (m,w) is an individually rational pair in the original market $(M,W,\succeq)$, thus Lemma 2 ensures that (m,w) is a blocking pair for $\tilde{\mu}$ in $(M,W,\succeq)$. Hence it holds

$$m \succ_w \tilde{\mu}(w) \text{ and } w \succ_m \tilde{\mu}(m). \tag{5}$$

But as $\tilde{\mu}$ is stable in $(M,W,\tilde{\succeq})$, it must hold that either $m \mathrel{\tilde{\precsim}_w} \tilde{\mu}(w)$ or $w \mathrel{\tilde{\precsim}_m} \tilde{\mu}(m)$. W.l.o.g. assume

$$m \mathrel{\tilde{\precsim}_w} \tilde{\mu}(w). \tag{6}$$

From $(m,w) \in \mathscr{B}(\mu)$ follows $m \in B_\mu(w)$ and thus $m \mathrel{\tilde{\succ}_w} w$ by (4). Therefore (6) implies $\tilde{\mu}(w) \neq w$, henceforth $(\tilde{\mu}(w),w) \in \tilde{\mu}$. With $\tilde{\mu} \subseteq \mathscr{B}(\mu)$ we conclude $(\tilde{\mu}(w),w) \in \mathscr{B}(\mu)$ and thus also $\tilde{\mu}(w) \in B_\mu(w)$.
We have shown that $m, \tilde{\mu}(w) \in B_\mu(w)$. But if this is true, from condition (3) and the first part of (5) follows

$$m \mathrel{\tilde{\succ}_w} \tilde{\mu}(w). \tag{7}$$

Clearly, (6) and (7) contradict each other. Herewith it is shown that every stable matching in the market $(M,W,\tilde{\succeq})$ is a permissible set of blocking pairs for μ in the market $(M,W,\succeq)$.
In the other direction, let D be a permissible set of blocking pairs for μ in the market $(M,W,\succeq)$. For notational convenience set $\mu_D := D$. $\mu_D \subseteq \mathscr{B}(\mu)$ and so for any pair $(m,w) \in \mu_D$ holds $w \in B_\mu(m)$ and $m \in B_\mu(w)$. Therefore condition (4) ensures $w \mathrel{\tilde{\succ}_m} m$ and $m \mathrel{\tilde{\succ}_w} w$, which means that μ_D is an individually rational matching in the market $(M,W,\tilde{\succeq})$. Now assume by contradiction that there exists a blocking pair $(\bar{m},\bar{w})$ for μ_D in the market $(M,W,\tilde{\succeq})$. Then $[\bar{m} \mathrel{\tilde{\succ}_{\bar{w}}} \mu_D(\bar{w}) \mathrel{\tilde{\succeq}_{\bar{w}}} \bar{w} \;\wedge\; \bar{w} \mathrel{\tilde{\succ}_{\bar{m}}} \mu_D(\bar{m}) \mathrel{\tilde{\succeq}_{\bar{m}}} \bar{m}]$, so by (4) follows $[\bar{m} \in B_\mu(\bar{w}) \;\wedge\; \bar{w} \in B_\mu(\bar{m})]$. It follows $(\bar{m},\bar{w}) \in \mathscr{B}(\mu)$ by the definitions of the sets $B_\mu(\bar{w})$ and $B_\mu(\bar{m})$. As μ_D is a permissible set of blocking pairs for μ in the market $(M,W,\succeq)$ and $(\bar{m},\bar{w}) \in \mathscr{B}(\mu)$, there must be a pair $(\hat{m},\hat{w}) \in \mu_D$ with $(\bar{m},\bar{w}) \in dom((\hat{m},\hat{w}))$, and it holds either $[\bar{m} = \hat{m} \;\wedge\; \hat{w} \succ_{\bar{m}} \bar{w}]$ or $[\bar{w} = \hat{w} \;\wedge\; \hat{m} \succ_{\bar{w}} \bar{m}]$. If $[\bar{m} = \hat{m} \;\wedge\; \hat{w} \succ_{\bar{m}} \bar{w}]$ is true, then $\hat{w} \in B_\mu(\bar{m})$ because $(\bar{m},\hat{w}) \in \mu_D \subseteq \mathscr{B}_\mu$. But if $\bar{w},\hat{w} \in B_\mu(\bar{m})$, from $\mu_D(\bar{m}) = \hat{w} \succ_{\bar{m}} \bar{w}$ and (3) follows $\mu_D(\bar{m}) = \hat{w} \mathrel{\tilde{\succ}_{\bar{m}}} \bar{w}$ which means that $(\bar{m},\bar{w})$ is no blocking pair for μ_D in the market $(M,W,\tilde{\succeq})$, contrary to our

assumption. If $[\bar{w} = \hat{w} \ \wedge \ \hat{m} \succ_{\bar{w}} \bar{m}]$ is true, the argument is symmetrical. $\square$

As a consequence, in order to find all permissible sets of blocking pairs for a matching μ, we just have to construct a preference profile $\tilde{\succeq}$ with regard to μ and then compute the set of stable matchings in the market $(M, W, \tilde{\succeq})$. Each of these matchings is a permissible set of blocking pairs for μ in the market $(M, W, \succeq)$.[28] An algorithm which can be applied to find the set of stable matchings in the market $(M, W, \tilde{\succeq})$ was devised by McVitie and Wilson (1971). If we are only interested in the cardinality of the permissible sets, of course it is sufficient to compute just one stable matching in the market $(M, W, \tilde{\succeq})$ with the algorithm of Gale and Shapley (1962).
We remark that Proposition 2 on page 12 could also be proved by making use of the well known result stated as Corollary 3 on page 20 together with Proposition 3 above. If all permissible sets of blocking pairs for a matching μ in a market $(M, W, \succeq)$ are in fact stable matchings in a manipulated market $(M, W, \tilde{\succeq})$ (Proposition 3), then the fact that the set of those players who are not single is the same at every stable matching in a market (Corollary 3) implies that all permissible sets of blocking pairs must be formed from the same set of players. However, by proving Proposition 2 independently of Corollary 3, it becomes clear that Proposition 2 directly follows from the Principles 1 and 2, through which permissible sets are defined, and does not depend on a hidden argument from the set of stable matchings.

7. A permissible set of blocking pairs for a matching μ is empty if and only if μ is stable

Now we come to the question under which condition nonemptiness of a permissible set is guaranteed. A permissible set $D(\mu) \subseteq \mathscr{B}(\mu)$ is an empty set if and only if μ is stable. Hence, according to proposition 1 there is no justified objection against a matching if and only if μ is stable. In particular, μ being a *weakly stable* matching (Klijn and Massó (2003)) is not sufficient for $D(\mu)$ to be empty. This is an important result, as it proves that the

[28] Our result reveals an isomorphism between the permissible sets of blocking pairs for a matching μ and the set of stable matchings in another marriage market. Similarly, Ehlers (2007) makes use of the result of Blair (1984) in order to show that there is an isomorphism between a Von Neumann-Morgenstern stable set in one marriage market and the set of stable matchings in another market (see Ehlers (2007), Remark 2, p. 544).

bargaining set presented in section 4 indeed coincides with the set of stable matchings.

Corollary 1 (to Proposition 3). *There is no justified objection against an individually rational matching μ if and only if μ is stable.*[29]

Proof. "$\Leftarrow$": If μ is stable, then there are no blocking pairs and thus there is no objection. "$\Rightarrow$": If μ is unstable, then let $(M, W, \tilde{\succeq})$ be a marriage market where $\tilde{\succeq}$ was constructed with regard to μ. In the first part of the proof of Proposition 3 was shown that any stable matching in $(M, W, \tilde{\succeq})$ is a nonempty permissible set of blocking pairs for μ, i.e. a justified objection against μ. $\square$

Corollary 2 (to Corollary 1). *If there exists an objection against a matching μ, then there exists a justified objection against μ.*

Proof. If there exists an objection S against μ, then any pair $(m, w) \in S$ is a blocking pair for μ. Thus μ is unstable. Hence, Corollary 1 ensures that there exists a justified objection against μ. $\square$

8. Permissible sets of blocking pairs and the set of stable matchings

Proposition 3 showed that for an arbitrary unstable matching μ in a marriage market $(M, W, \succeq)$ we can construct another marriage market $(M, W, \tilde{\succeq})$ such that the stable matchings in $(M, W, \tilde{\succeq})$ are the permissible sets of blocking pairs for μ. Can we reverse the direction of this argument? If we have an arbitrary market $(M, W, \succeq)$, can we always find a matching μ such that the set of stable matchings of $(M, W, \succeq)$ are the permissible sets of blocking pairs for μ? In this section that question is answered affirmatively. Surprisingly, the stable matchings in an arbitrary market $(M, W, \succeq)$ are the permissible sets of blocking pairs for the *empty matching* (the matching in which all players are single) of $(M, W, \succeq)$:

Proposition 4. *Let μ be the empty matching in the market $(M, W, \succeq)$. A matching μ' is stable in $(M, W, \succeq)$ if and only if it is a permissible set of blocking pairs for the matching μ.*

[29] Remind that by definition an objection is nonempty. Thus empty justified objections do not exist.

Proof. First assume μ' is a stable matching in $(M, W, \succeq)$. Let (m, w) be a blocking pair for the empty matching such that (m, w) is not an element of μ'. To establish that μ' is a permissible set of blocking pairs for μ, it must be shown that (m, w) is dominated by a pair in μ'. (m, w) cannot be a blocking pair for μ' due to the stability of μ'. So because of the strict preferences, it must hold $\mu'(m) \succ_m w$ or $\mu'(w) \succ_w m$ (or both), as otherwise (m, w) would block μ'. But then (m, w) is dominated either by $(m, \mu'(m)) \in \mu'$ or by $(\mu'(w), w) \in \mu'$. In the other direction, assume D is a permissible set of blocking pairs for μ. Then D is a matching in $(M, W, \succeq)$ (with all players who are not part of a pair in D being singles), and D is individually rational because $D \subseteq \mathscr{B}(\mu)$.[30] It needs to be shown that there exist no blocking pairs for the matching D. A blocking pair (m, w) for D cannot be dominated by any pair in D because then either m or w would prefer their partners under D over the formation of (m, w) and so (m, w) would be no blocking pair. But if (m, w) is not dominated by any pair in D, then it is a counterobjection against D. By Proposition 1 D is a justified objection against μ, and thus there exist no counterobjections against D. This establishes that there are no blocking pairs for D. $\square$

Using Propositions 4 and 2, a well known result of matching theory can be proved again:

Corollary 3. *Let μ, μ' be stable matchings in a market $(M, W, \succeq)$. Then a player $x \in M \cup W$ who is not single at the matching μ is also not single at the matching μ'.*

Proof. By Proposition 4 both μ and μ' are permissible sets of blocking pairs for the empty matching in the market $(M, W, \succeq)$, and thus justified objections against that matching. So from Proposition 2 follows $p(\mu) = p(\mu')$. $\square$

This result was proved in two different ways for marriage markets in McVitie and Wilson (1970) (theorem on page 298) and Gale and Sotomayor (1985) (Proposition 1). It was proved in Roth (1984) (Theorem 9) for the generalization of many-to-one matching problems.

[30] Under the empty matching μ, all players are singles. So for the members of any blocking pair (m, w) for μ holds $m \succ_w \mu(w) = w$ and $w \succ_m \mu(m) = m$. Hence from $D \subseteq \mathscr{B}(\mu)$ follows the individual rationality of D.

9. Comparing the cardinalities of the sets $D(\mu)$ and $\mathscr{B}(\mu)$

The following example demonstrates that the new measure can be reversed to the measures of instability frequently used in previous works, namely the numbers of blocking pairs.

Example: Consider a market with

$$M = \{m_1, m_2, m_3\} \text{ and } W = \{w_1, w_2, w_3, w_4, w_5\}$$

and the following preferences:[31]

$$\begin{aligned} P(w_1) &= m_1 \succ_{w_1} w_1 & \quad P(m_1) &= w_1 \succ_{m_1} w_2 \succ_{m_1} w_3 \succ_{m_1} m_1 \\ P(w_2) &= m_1 \succ_{w_2} w_2 & P(m_2) &= w_4 \succ_{m_2} m_2 \\ P(w_3) &= m_1 \succ_{w_3} w_3 & P(m_3) &= w_5 \succ_{m_3} m_3 \\ P(w_4) &= m_2 \succ_{w_4} w_4 & & \\ P(w_5) &= m_3 \succ_{w_5} w_5 & & \end{aligned}$$

Compare two matchings μ_1, μ_2 in this market, defined by

$$\begin{aligned} \mu_1 &= \{(m_2, w_4), (m_3, w_5)\}, \\ \mu_2 &= \{(m_1, w_1)\}. \end{aligned}$$

Then $\mathscr{B}(\mu_1) = \{(m_1, w_1), (m_1, w_2), (m_1, w_3)\}$ and $\mathscr{B}(\mu_2) = \{(m_2, w_4), (m_3, w_5)\}$. As $|\mathscr{B}(\mu_1)| = 3 > 2 = |\mathscr{B}(\mu_2)|$, the matching μ_1 would be considered less stable than μ_2 by the traditional measure. In constrast, we have $D(\mu_1) = \{(m_1, w_1)\}$ and $D(\mu_2) = \{(m_2, w_4), (m_3, w_5)\}$. So $|D(\mu_1)| = 1 < 2 = |D(\mu_2)|$. By the measure proposed here, matching μ_2 would be less stable than μ_1.

So the new measure and the number of blocking pairs can be reversed to each other. But to what extent can they be reversed? It would be at odds with intuition if one could find two matchings μ and μ' in the same market with $\mathscr{B}(\mu) \subseteq \mathscr{B}(\mu')$ and $|D(\mu)| > |D(\mu')|$. In that case a manipulation

[31] The preferences are only stated down to the single option. Partners who are less preferred than the single option are irrelevant because both matchings in the example are individually rational.

of the matching μ which would *add* blocking pairs to $\mathscr{B}(\mu)$ could lead to a *reduction* of the measure. If such matchings μ and μ' existed, one could hardly claim that the cardinality of a permissible set is a reasonable estimate for the "degree of instability" of a matching. Fortunately, this possibility can be ruled out:

Proposition 5. *Let μ and μ' be individually rational matchings with $\mathscr{B}(\mu) \subseteq \mathscr{B}(\mu')$. Then $|D(\mu)| \leq |D(\mu')|$.*

Proof. An injective function which maps $D(\mu)$ into $D(\mu')$ will be constructed. Such a function only exists if $|D(\mu)| \leq |D(\mu')|$. For $x \in D(\mu)$, let $y(x) := (m, w)$ be an arbitrary element in $D(\mu')$ with $x \in dom((m, w))$. Define a function $f : D(\mu) \longrightarrow D(\mu')$ by

$$f(x) = \begin{cases} x & \text{if } x \in D(\mu) \cap D(\mu'), \\ y(x) & \text{otherwise.} \end{cases}$$

It will be shown that the function $f(x)$ exists and that it is an injection. $f(x)$ exists because any element $x \in D(\mu)$ which is not in $D(\mu')$ must be dominated by an element in $D(\mu')$ (Principle 2). So $y(x)$ exists for each $x \in D(\mu)$, $x \notin D(\mu')$. Furthermore, for any $x \notin D(\mu) \cap D(\mu')$ holds $y(x) \notin D(\mu) \cap D(\mu')$. This follows from principle 1: $y(x)$ is a pair which dominates x, which means that x and $y(x)$ must have a player in common. But as $D(\mu)$ is a matching (principle 1), no two pairs in the set $D(\mu)$ share a player.
Hence, the set $D(\mu) \backslash D(\mu')$ is mapped into $D(\mu') \backslash D(\mu)$. Consequently, f is an injection on the domain $D(\mu) \cap D(\mu')$.
Now assume there would be two pairs $x, z \in D(\mu) \backslash D(\mu')$ with $y(x) = y(z) := (m, w)$. Note that one pair in $D(\mu')$ cannot dominate *two* pairs in $D(\mu)$ via m, because then m would be a member of *both* these dominated pairs. But $D(\mu)$ is a matching, and there are no two pairs which share a player. (m can be replaced by w in this argument.) Therefore x is dominated by (m, w) via m and z is dominated by (m, w) via w (or vice versa). But if this is the case, m and w prefer the pair (m, w) over the pairs x and z, whence the pair (m, w) is a blocking pair in $\mathscr{B}(\mu)$. So (m, w), which is not member of $D(\mu)$, must be dominated by a pair in $D(\mu)$ (Principle 2). But m and w prefer (m, w) over x and z, implying (m, w) can neither be dominated by x nor by z. Hence, (m, w) is not dominated by any pair in $D(\mu)$, contradicting Principle 2. This rules out $y(x) = y(z)$. It follows that f is also an injection on the domain $D(\mu) \backslash D(\mu')$. $\square$

10. Conclusion

A measure to compare matchings is needed in situations in which Gale-Shapley stability is not a feasible or appropriate objective. Such situations can emerge in experimental economics, when the matchmaker pursues goals other than stability, or when particular market circumstances prevent the matchmaker from generating a stable matching. The number of blocking pairs or closely related criteria were deployed in previous papers (Khuller et al. (1994), Roth and Xing (1997), Niederle and Roth (2007), Hamada et al. (2009), Biró et al. (2010)), and even some general properties on these measures of instability were derived in Eriksson and Häggström (2008).
Here we argued that instead of counting all blocking pairs which exist for a matching, one should rather count a set of blocking pairs which comprises a possible and economically reasonable transformation of the market. This approach led to two principles for the set of blocking pairs to be counted: The first made sure that this set was really a matching, which means that it must not contain two or more pairs sharing a common member. The second principle was based on economic intuition. It stated that any blocking pair (m, w) for a matching μ was counted, unless one counted another blocking pair $(\bar{m}, \bar{w})$ for μ which *dominated* (m, w). The economic argument behind this principle is that the sole reason a blocking pair would not get satisfied should be the existence of another, dominating blocking pair which *does* get satisfied. Those sets of blocking pairs which fulfilled both principles were called *permissible*.
It was then shown that we can characterize the set of stable matchings as a bargaining set, and that according to this bargaining set the permissible sets of blocking pairs are justified objections of groups of players against the given matching. Furthermore, even if multiple different permissible sets of blocking pairs exist for a matching, they all have the same cardinality. This property makes the cardinality of permissible sets a practicable measure of matching comparison. Next, a method was presented to identify all permissible sets of blocking pairs existing for a matching. Then we established that there is no unstable matching which has an empty permissible set of blocking pairs, and no stable matching which has a non-empty permissible set of blocking pairs. Finally, an example illustrated that for two matchings in the same market, the measure brought forward in this paper can be conversed to the absolute and relative numbers of blocking pairs.
In this theoretical work, one important question remained unanswered: Does

the cardinality of a permissible set of blocking pairs *empirically* capture the extent of partner changes inherent in an unstable matching?
In order to substantiate such a claim, one could conduct a laboratory experiment and check whether the *first* blocking pairs which get satisfied at a given unstable matching are sufficiently often permissible sets. Yet even if the relevance of permissible sets could be empirically supported, their cardinality would just predict the extent of the *first stage* of the dynamic.[32]
Divorce cost may prevent excessive partner changes and cause dynamics to be short. Therefore we argued that the costlier divorces are, the more adequate it is to measure the "degree of instability" by the size of permissible sets. However, in real world matching markets, the average number of players' partner changes may be easier observable than divorce costs (which are not necessarily monetary).[33] The lower the average number of partner changes is, the higher tends to be the share of the market readjustment occuring at the beginning of the dynamic, and the more relevant is the size of the permissible sets. Consequently, real world markets with low average numbers of partner changes would be the first candidates for applying the measure introduced in this article.

Appendix A. Proofs left out in the main body

Lemma 1 (of page 9). *Let (m,w) and $(\hat{m},\hat{w})$ be two different blocking pairs in $\mathscr{B}(\mu)$. If $m=\hat{m}$ or $w=\hat{w}$, then exactly one of the following statements is true:*

1. $(m,w)\in dom((\hat{m},\hat{w}))$
2. $(\hat{m},\hat{w})\in dom((m,w))$.

Proof. Only the case $m=\hat{m}$ is considered, as the case $w=\hat{w}$ can be treated analogously. First we prove that at least one of the two cases holds. Assume $(m,w)\notin dom((\hat{m},\hat{w}))$. Then $m=\hat{m}$ implies $\hat{w}\not\succ_m w$, and thus the strict preferences demand $w\succ_m\hat{w}$, which means $(\hat{m},\hat{w})\in dom((m,w))$. By the same argument, $(\hat{m},\hat{w})\notin dom((m,w))$ implies $(m,w)\in dom((\hat{m},\hat{w}))$.

[32] See Appendix B for a rudimentary analysis of the n^{th}-order dynamics.

[33] For *real* marriage markets (formed by men and women who want to marry), data about the number of partner changes, divorce rates, and players' search efforts were acquired by sociologists. Frey and Eichenberger (1996) highlight economically interesting facts from this literature.

Now we prove that $(m, w) \in dom((\hat{m}, \hat{w}))$ and $(\hat{m}, \hat{w}) \in dom((m, w))$ cannot be true at the same time. Assume $(m, w) \in dom((\hat{m}, \hat{w}))$. The definition of $(m, w) \in dom((\hat{m}, \hat{w}))$ requires that either holds 1) $[m = \hat{m}$ and $\hat{w} \succ_m w]$ or 2) $[w = \hat{w}$ and $\hat{m} \succ_w m]$. From $m = \hat{m}$ follows $w \neq \hat{w}$ because of $(m, w) \neq (\hat{m}, \hat{w})$, and so not both cases can be fulfilled simultaneously. Thus only case 1) holds true. $m = \hat{m}$ and $(\hat{m}, \hat{w}) \in dom((m, w))$ would imply $w \succ_m \hat{w}$, a contradiction. Hence it follows $(\hat{m}, \hat{w}) \notin dom((m, w))$. $\square$

Proposition 1 (of page 11). *$D(\mu)$ is a permissible set of blocking pairs if and only if it is a justified objection against μ.*

Proof. Assume that the set $D(\mu)$ is a justified objection against μ. Then definition 7 ensures that $D(\mu)$ is a matching (i.e. fulfills Principle 1) and $D(\mu) \subseteq \mathscr{B}(\mu)$. It needs to be shown that $D(\mu)$ also satisfies Principle 2. By contradiction, if $D(\mu)$ would not satisfy Principle 2, there would be a pair $(m, w) \in \mathscr{B}(\mu) \backslash D(\mu)$ for which exists no pair $(\hat{m}, \hat{w}) \in D(\mu)$ with $(m, w) \in dom((\hat{m}, \hat{w}))$. But if $(m, w) \notin dom((\hat{m}, \hat{w}))$ for any $(\hat{m}, \hat{w}) \in D(\mu)$, the pair (m, w) fits the definition of a counterobjection, generating a contradiction. Hence $D(\mu)$ must also fulfill Principle 2, whence it follows that $D(\mu)$ is a permissible set of blocking pairs. In the other direction, assume $D(\mu)$ is a permissible set of blocking pairs but there is a counterobjection $T \subseteq \mathscr{B}(\mu)$ against $D(\mu)$. From this we construct a contradiction as follows: Pick a pair $(m, w) \in T$, $(m, w) \notin D(\mu)$. Such a pair must exist since $T \nsubseteq D(\mu)$ is ensured by the definition of a counterobjection. Also by the definition of a counterobjection, no pair $(\hat{m}, \hat{w}) \in D(\mu)$ dominates (m, w). But as $(m, w) \in \mathscr{B}(\mu) \backslash D(\mu)$, Principle 2 demands that there is a pair in $D(\mu)$ which dominates (m, w), delivering the contradiction. $\square$

Lemma 2 (of page 16). *Let μ be a matching in a market $(M, W, \succeq)$ and let $(m, w) \in M \times W$ be an individually rational pair*[34] *with $(m, w) \notin \mu$. If there is no $(\hat{m}, \hat{w}) \in \mu$ with $(m, w) \in dom(\hat{m}, \hat{w})$, then (m, w) is a blocking pair for μ.*

Proof. Assume by contradiction that an individually rational pair $(m, w) \in M \times W$ with $(m, w) \notin \mu$ would neither be dominated by any pair in μ, *nor* would it be be a blocking pair for μ. As (m, w) is no blocking pair, it must

[34] This means that both $w \succ_m m$ and $m \succ_w w$ are fulfilled.

hold $\mu(m) \succ_m w$ or $\mu(w) \succ_w m$, implying $\mu(m) \neq m$ or $\mu(w) \neq w$ because (m, w) is an individually rational pair. For any $x \in M \cup W$ and any matching μ holds $\mu(x) \neq x \Rightarrow (x, \mu(x)) \in \mu \vee (\mu(x), x) \in \mu$. Hence at least one of the pairs $(m, \mu(m)), (\mu(w), w)$ is an element of μ. As a consequence, it holds $(m, w) \in dom((m, \mu(m)))$ or $(m, w) \in dom((\mu(w), w))$, contradicting the assumption that (m, w) was not dominated by any pair in μ. $\square$

Appendix B. n^{th}-Order Dynamics

Iteratively identifying permissible sets and satisfying their blocking pairs leads to a dynamic of matching transformations. It will be shown that the dynamic does not necessarily converge to a stable matching. Therefore the size of a permissible set $D(\mu)$ cannot be interpreted as a "distance from the set of stable matchings" in a direct sense, i.e. as the length of a sequence which transforms μ into a stable matching. Besides this fact, extending the analysis to multiple stages of matching transformation yields no insights relevant for the measure discussed in the preceding sections. This is the reason why we refer to this topic in the appendix.
n^{th}-order dynamics will be just touched here without comprehensive analysis. Nevertheless, the example presented in this section gives rise to some questions which might inspire future research.

Let μ be a matching and $D(\mu)$ be a permissible set of blocking pairs for μ. A new matching $\sigma(\mu)$ can be defined by

$$\sigma(\mu) := D(\mu) \cup \{(m, w) \in \mu \ \mid \ m, w \notin p(D(\mu))\}$$

with $p(D(\mu))$ being set of players who are not singles under the matching $D(\mu)$ (see the definition on page 8). Verbally, the matching $\sigma(\mu)$ contains the permissible set $D(\mu)$ and those pairs of μ which do not share a player with any pair in $D(\mu)$. The definition of $\sigma(\mu)$ implies that a player x who does not participate in the permissible set $D(\mu)$, but whose *partner* $\mu(x)$ does, is single under $\sigma(\mu)$.[35] Using the transformation σ, one can think about a sequence of matchings $(\mu^i)_{i=0,\ldots,\infty}$ such that $\mu^i = \sigma(\mu^{i-1})$ for some permissible

[35] For satisfying arbitrary single blocking pairs (not permissible sets), this natural way to transform a matching was formalized in Blum and Rothblum (2002), p. 432. Furthermore, the matching transformation defined here is a special case of the *enforceability notion* defined by Klaus et al. (forthcoming) for roommate problems, where only coalitions are considered in which all players improve their outcome.

set $D(\mu^{i-1})$.

By Corollary 1 (page 19), if the sequence $(\mu^i)_{i=1,\dots,\infty}$ converges, then it must converge to a stable matching. Unfortunately, it may happen that the sequence cycles, and even worse, there may not even be a "way out" of the cycle through choosing a different permissible set of blocking pairs at some point. This fact is illustrated by the upcoming example. To understand the example correctly, the following (trivial) result is useful:

Lemma 3. *Let μ be a matching. If $\mathscr{B}(\mu)$ is a matching, then there exists a unique permissible set of blocking pairs for μ.*

Proof. If $\mathscr{B}(\mu)$ is a matching, then no two pairs in $\mathscr{B}(\mu)$ share a player, hence no pair in $\mathscr{B}(\mu)$ is dominated by any other pair in $\mathscr{B}(\mu)$. Therefore by principle 2 all pairs in $\mathscr{B}(\mu)$ must be in $D(\mu)$, i.e. $\mathscr{B}(\mu) \subseteq D(\mu)$. By definition holds $D(\mu) \subseteq \mathscr{B}(\mu)$ (cf. page 8), thus for all permissible sets $D(\mu)$ for μ must hold $\mathscr{B}(\mu) = D(\mu)$, whence $D(\mu)$ is a unique set. □

Example: Consider a market with

$$M = \{m_1, m_2, m_3\} \text{ and } W = \{w_1, w_2, w_3\}$$

and the following preferences:[36]

$$\begin{aligned} P(w_1) &= \ m_3 \succ_{w_1} m_1 \succ_{w_1} w_1 & \qquad P(m_1) &= w_1 \succ_{m_1} w_3 \succ_{m_1} m_1 \\ P(w_2) &= \ m_2 \succ_{w_2} m_3 \succ_{w_2} w_2 & \qquad P(m_2) &= w_3 \succ_{m_2} w_2 \succ_{m_2} m_2 \\ P(w_3) &= \ m_1 \succ_{w_3} m_2 \succ_{w_3} w_3 & \qquad P(m_3) &= w_2 \succ_{m_3} w_1 \succ_{m_3} m_3 \end{aligned}$$

The cycle starts with the matching μ^0 in this market, defined by

$$\mu^0 = \{(m_1, w_1), (m_2, w_2)\}.$$

Then $\mathscr{B}(\mu^0) = \{(m_3, w_1), (m_2, w_3)\}$ (the set of blocking pairs for μ^0). Because $\mathscr{B}(\mu^0)$ is a matching, Lemma 3 ensures $D(\mu^0) = \mathscr{B}(\mu^0)$ and $D(\mu^0)$ is unique. By the transformation σ we obtain

$$\mu^1 = \{(m_3, w_1), (m_2, w_3)\}.$$

[36] As before, the preferences are only stated down to the single option.

Again $\mathscr{B}(\mu^1) = \{(m_1, w_3), (m_3, w_2)\}$ is a matching, so by lemma 3 follows $D(\mu^1) = \mathscr{B}(\mu^1)$. Hence

$$\mu^2 = \{(m_1, w_3), (m_3, w_2)\}.$$

Again $\mathscr{B}(\mu^2) = \{(m_1, w_1), (m_2, w_2)\}$ is a matching, so $\mathscr{B}(\mu^2) = D(\mu^2)$. Therefore

$$\mu^3 = \{(m_1, w_1), (m_2, w_2)\}$$

and $\mu^0 = \mu^3$.

At each step of the transformation, *all* existing blocking pairs were satisfied simultaneously. So the example demonstrates that stability cannot always be reached if one satisfies more than one blocking pair at each matching. This is an interesting fact, given a sequence which leads to stability if one chooses just one blocking pair at a matching *always* exists (Roth and Vande Vate (1990)). How do those matchings which can be reached through any of the sequences starting at μ^0 depend on μ^0? If one could prove nice properties of the sequences defined in this section, and empirical support could be delivered, they might be promising candidates for a genuine model of matching market microdynamics.

References

Biró, P., Manlove, D.F., Mittal, S., 2010. Size versus Stability in the Marriage Problem. Theoretical Computer Science 411, 1828–1841.

Blair, C., 1984. Every finite distributive Lattice is a Set of stable Matchings. Journal of Combinatorial Theory (Series A) 37, 353–356.

Blum, Y., Rothblum, U.G., 2002. 'Timing is Everything' and Marital Bliss. Journal of Economic Theory 103, 429–443.

Cechlárová, K., 2002. Randomized Matching Mechanism revisited. Preprint.

Diamantoudi, E., Miyagawa, E., Xue, L., 2004. Random Paths to Stability in the Roommate Problem. Games and Economic Behavior 48, 18–28.

Ehlers, L., 2007. Von Neumann-Morgenstern Stable Sets in Matching Problems. Journal of Economic Theory 134, 537–547.

Eriksson, K., Häggström, O., 2008. Instability of Matchings in decentralized Markets with various Preference Structures. International Journal of Game Theory 36, 409–420.

Frey, B.S., Eichenberger, R., 1996. Marriage Paradoxes. Rationality and Society 8, 187–206.

Gale, D., Shapley, L.S., 1962. College Admissions and the Stability of Marriage. American Mathematical Monthly 69, 9–15.

Gale, D., Sotomayor, M., 1985. Ms. Machiavelli and the Stable Matching Problem. American Mathematical Monthly 92, 261–286.

Hamada, K., Iwama, K., Miyazaki, S., 2009. An improved Approximation Lower Bound for finding almost stable Maximum Matchings. Information Processing Letters 109, 1036–1040.

Kahan, J.P., Rapoport, A., 1984. Theories of Coalition Formation. L. Erlbaum Associates.

Khuller, S., Mitchell, S.G., Vazirani, V.V., 1994. On-line Algorithms for weighted bipartite Matching and Stable Marriages. Theoretical Computer Science 127, 255–267.

Klaus, B., Klijn, F., 2010. Smith and Rawls share a Room: Stability and Medians. Social Choice and Welfare 35, 647–667.

Klaus, B., Klijn, F., Walzl, M., 2009. Farsighted Stability for Roommate Markets. Harvard Business School Working Paper 09-135 (`www.hbs.edu/research/pdf/09-135.pdf`).

Klaus, B., Klijn, F., Walzl, M., forthcoming. Farsighted Stability for Roommate Markets. Journal of Public Economic Theory .

Klijn, F., Massó, J., 2003. Weak Stability and a Bargaining set for the Marriage Model. Games and Economic Behavior 42, 91–100.

Mas-Colell, A., 1989. An Equivalence Theorem for a Bargaining Set. Journal of Mathematical Economics 18, 129–139.

Maschler, M., 1992. The Bargaining set, Kernel, and Nucleolus, in: Aumann, R.J., Hart, S. (Eds.), Handbook of Game Theory, Vol. 1. Elsevier, pp. 591–667.

McVitie, D.G., Wilson, L.B., 1970. Stable Marriage Assignments for Unequal Sets. BIT 10, 295–309.

McVitie, D.G., Wilson, L.B., 1971. The Stable Marriage Problem. Communications of the ACM 14, 486–492.

Niederle, M., Roth, A.E., 2003. Unraveling reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match. Journal of Political Economy 111, 1342–1352.

Niederle, M., Roth, A.E., 2007. Making Markets Thick: Designing Rules for Offers and Acceptances. Preprint.

Peleg, B., Sudhölter, P., 2007. Introduction to the Theory of Cooperative Games. Springer-Verlag. 2nd edition.

Roberts, F.S., Tesman, B., 2009. Applied Combinatorics. CRC Press. 2nd edition.

Roth, A.E., 1984. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. Journal of Political Economy 92, 991–1016.

Roth, A.E., 1991. A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K. American Economic Review 81, 415–440.

Roth, A.E., Sotomayor, M., 1990. Two-Sided Matching - A Study in Game Theoretic Modeling and Analysis. Cambridge University Press.

Roth, A.E., Vande Vate, J.H., 1990. Random Paths to Stability in Two-Sided Matching. Econometrica 58, 1475–1480.

Roth, A.E., Xing, X., 1997. Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in the Market for Clinical Psychologists. Journal of Political Economy 105, 284–329.

Zhou, L., 1994. A New Bargaining Set of an N-Person Game and Endogenous Coalition Formation. Games and Economic Behavior 6, 512–526.